Name: Md. Abdul Mukit

Class Roll: 2407

Exam Roll: 170491

Course Name: Machine Learning Lab

Course Code: CSE-458

Experiment Name: Weka for Machine Learning

Submitted To

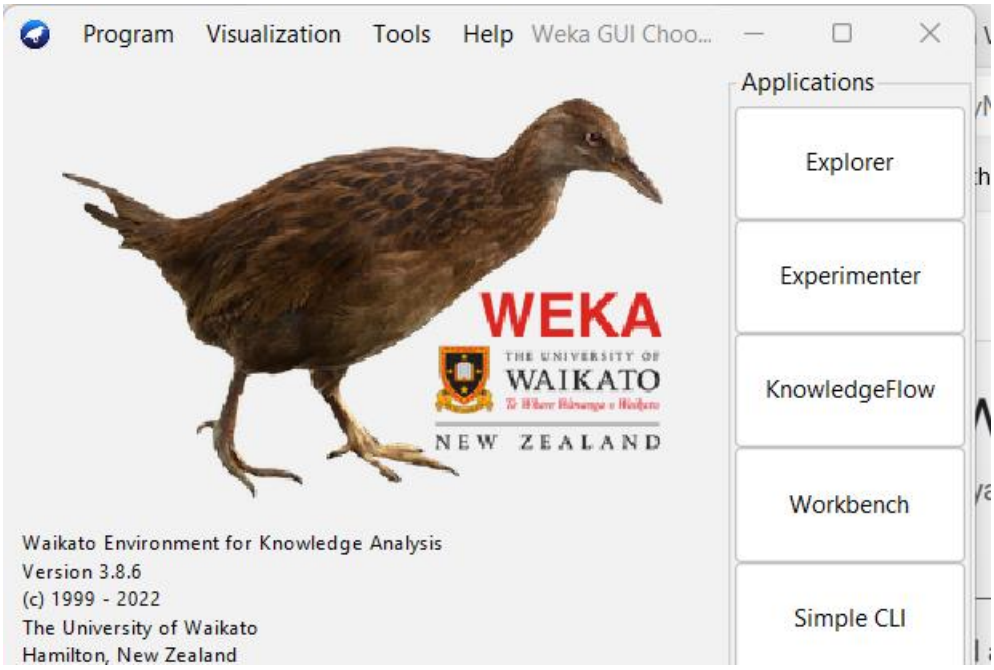
Sarnali Basak

Associate Prof. Dept of CSE, JU

Md Masum Bhuyan

Lecturer Dept of CSE, JU

In this lab experiment we explore weka for machine learning purpose. After downloading weka from wekato we installed it. Then we open weka in explorer category.



Step2: we load iris dataset from weka dataset. And visualize dataset. Iris Dataset use 4 parameter sepallength, sepalwidth, petallength and petalwidth. Based on the parameter value weka classify an object into three category

- Iris-setosa
- Iris-versicolor
- Iris-virginica

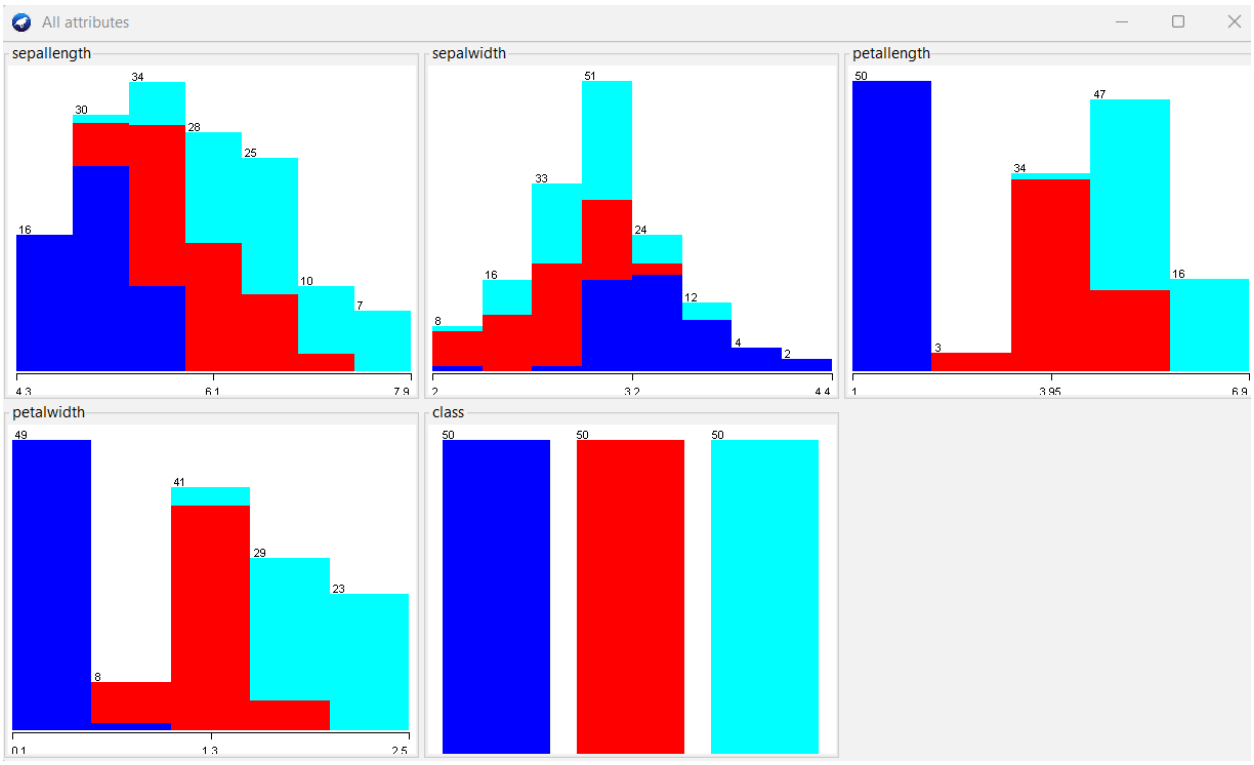


Fig: Visualize Iris dataset

Step-3: Visualize dataset

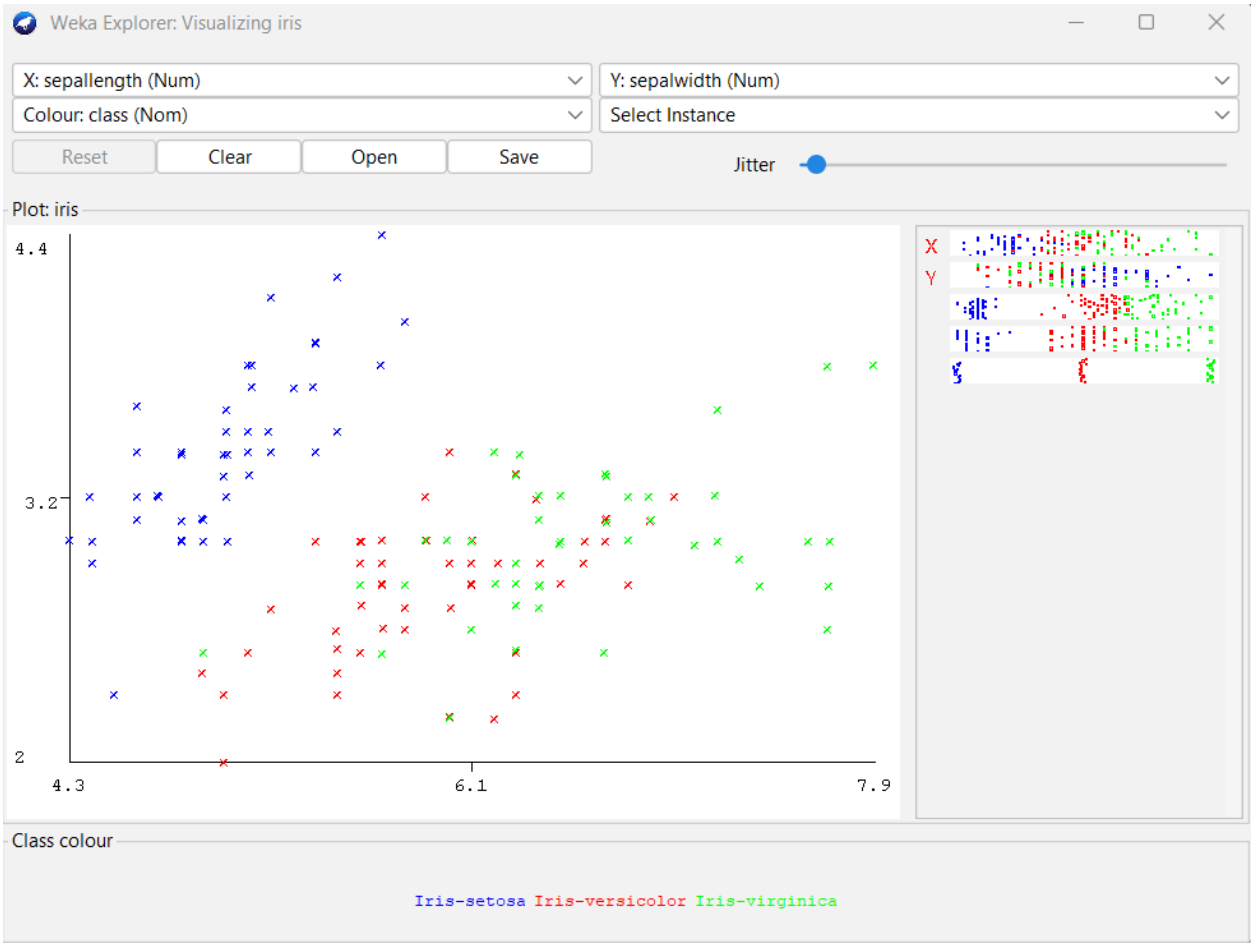
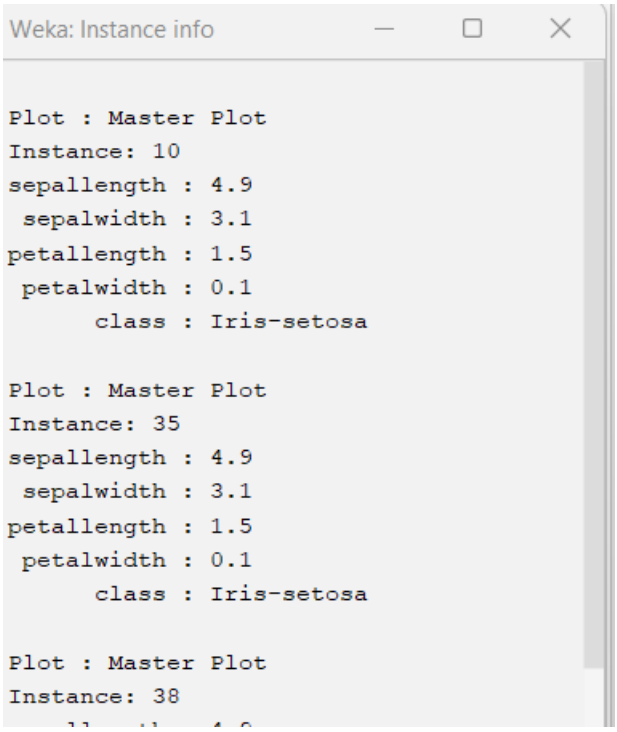


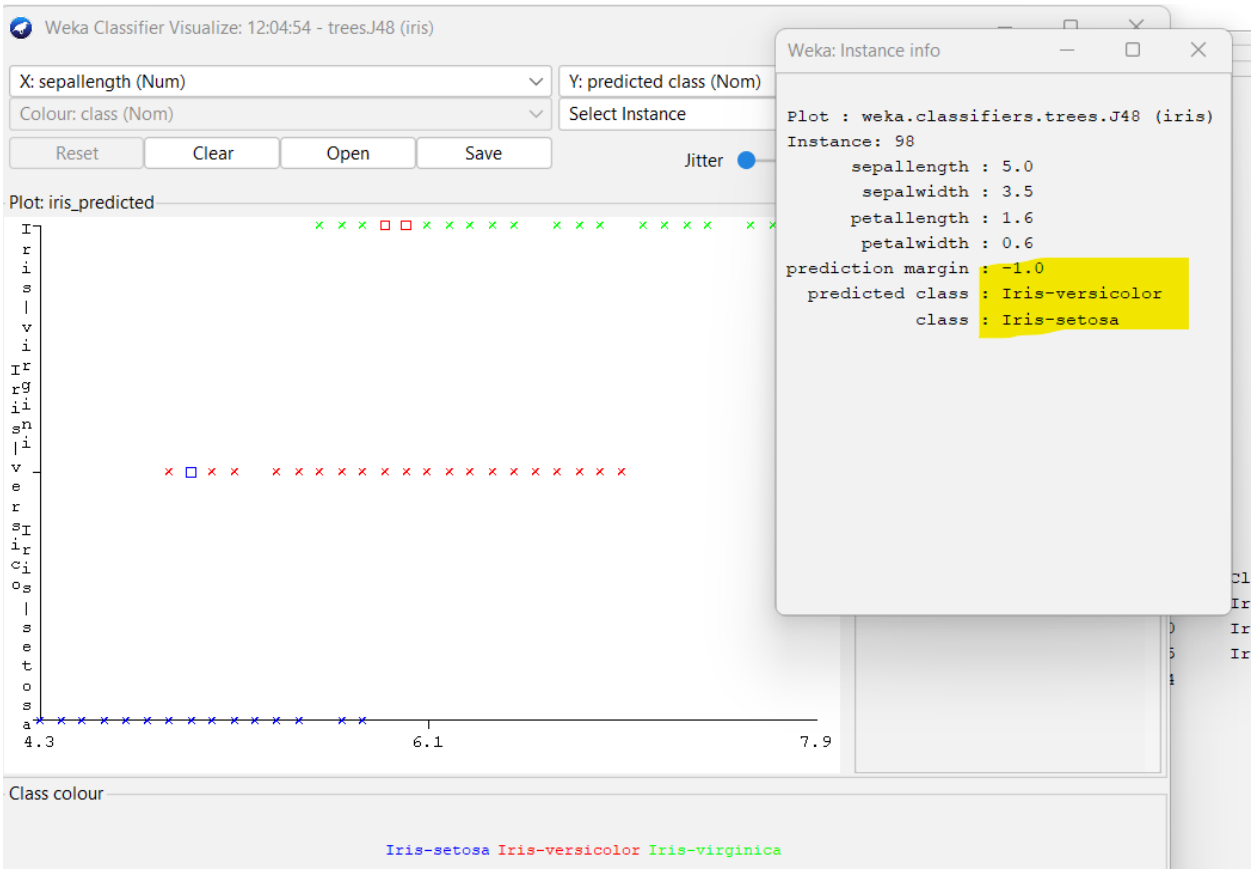
Fig: Data plotting based on sapal-length and sepal-width

After modifying jitter plot we can visualize each enstance.



Visualize classification error.

After using classification using j48 algorithm. We right click buffer and select visualize classification error and inspect data. Here – the predicted class is Iris-versicolor where actual lebeled class is Iris-setosa.

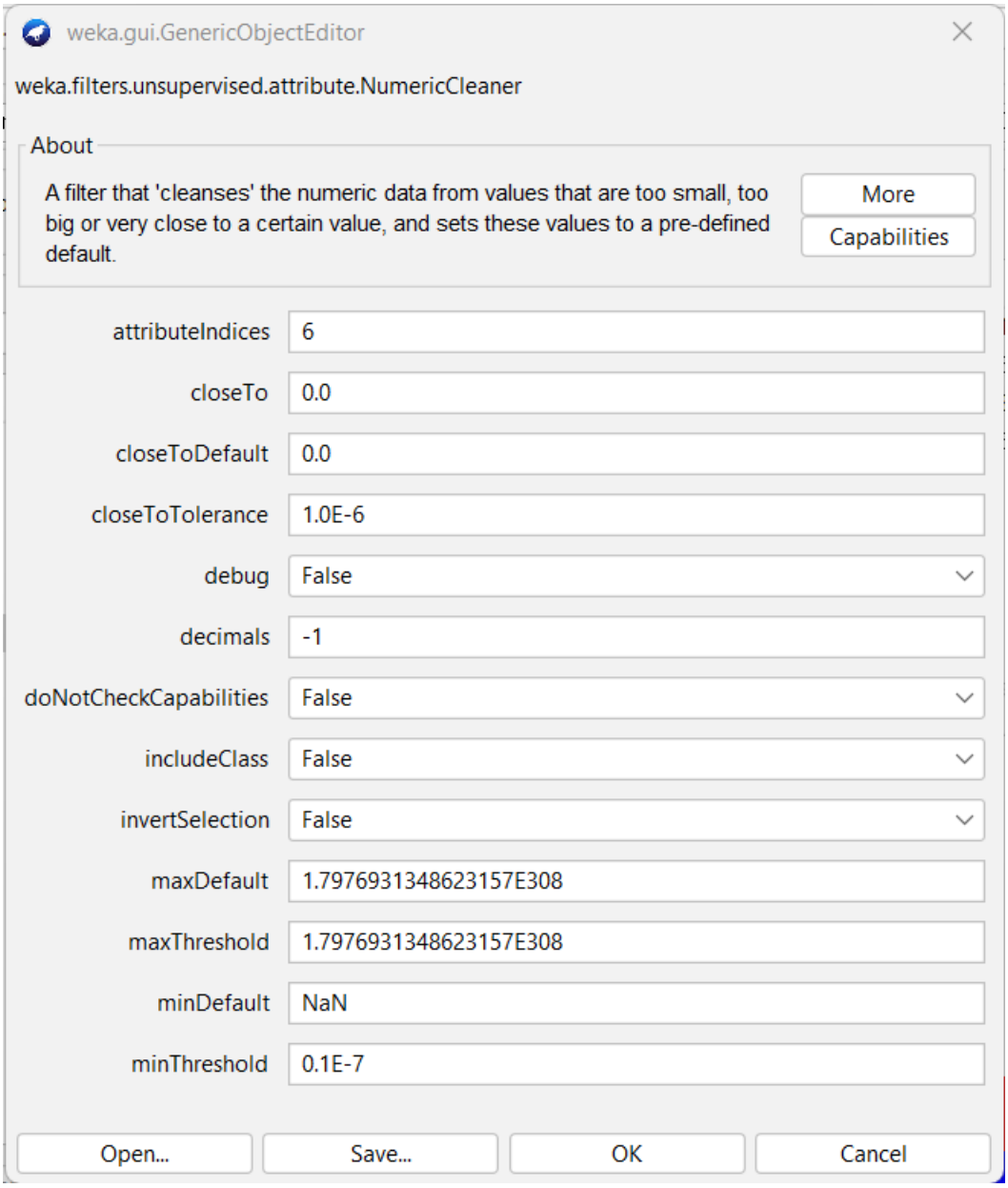


Step-4: Handling missing value

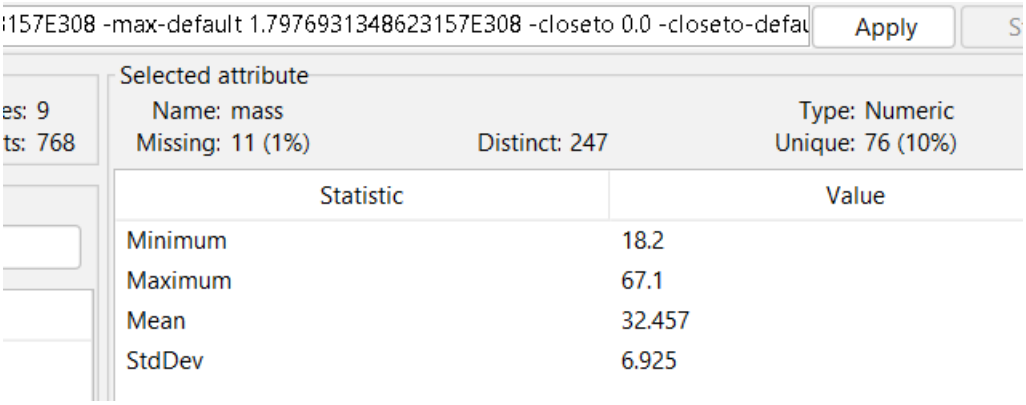
To perform these we choose weka diabetes dataset. These dataset predict instance diabetes positive or negative based on 8 parameter. We check missing value from **mass** index-6. To do this on preprocess tab we select

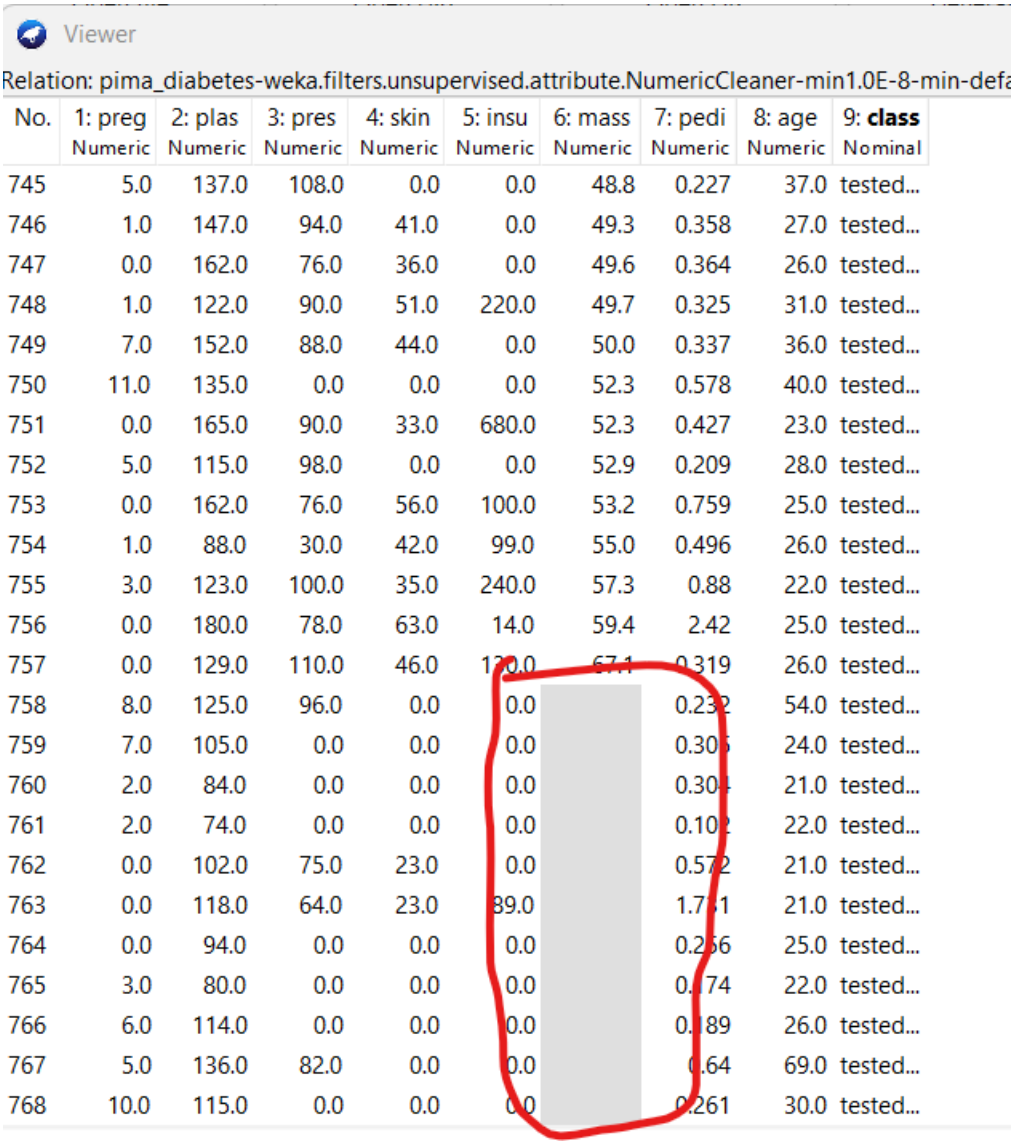
Choose > Filter > Unsupervised > attribute > NumericCleaner

In the filter section we select MinDefault value to NaN. Min threshold to 1E-8 and attributeIndices-6 as shown in figure. Then click apply.



Then we have 11 missing values in mass section.





No.	1: preg Numeric	2: plas Numeric	3: pres Numeric	4: skin Numeric	5: insu Numeric	6: mass Numeric	7: pedi Numeric	8: age Numeric	9: **class** Nominal
745	5.0	137.0	108.0	0.0	0.0	48.8	0.227	37.0	tested...
746	1.0	147.0	94.0	41.0	0.0	49.3	0.358	27.0	tested...
747	0.0	162.0	76.0	36.0	0.0	49.6	0.364	26.0	tested...
748	1.0	122.0	90.0	51.0	220.0	49.7	0.325	31.0	tested...
749	7.0	152.0	88.0	44.0	0.0	50.0	0.337	36.0	tested...
750	11.0	135.0	0.0	0.0	0.0	52.3	0.578	40.0	tested...
751	0.0	165.0	90.0	33.0	680.0	52.3	0.427	23.0	tested...
752	5.0	115.0	98.0	0.0	0.0	52.9	0.209	28.0	tested...
753	0.0	162.0	76.0	56.0	100.0	53.2	0.759	25.0	tested...
754	1.0	88.0	30.0	42.0	99.0	55.0	0.496	26.0	tested...
755	3.0	123.0	100.0	35.0	240.0	57.3	0.88	22.0	tested...
756	0.0	180.0	78.0	63.0	14.0	59.4	2.42	25.0	tested...
757	0.0	129.0	110.0	46.0	130.0	67.1	0.319	26.0	tested...
758	8.0	125.0	96.0	0.0	0.0		0.232	54.0	tested...
759	7.0	105.0	0.0	0.0	0.0		0.305	24.0	tested...
760	2.0	84.0	0.0	0.0	0.0		0.304	21.0	tested...
761	2.0	74.0	0.0	0.0	0.0		0.102	22.0	tested...
762	0.0	102.0	75.0	23.0	0.0		0.572	21.0	tested...
763	0.0	118.0	64.0	23.0	89.0		1.731	21.0	tested...
764	0.0	94.0	0.0	0.0	0.0		0.256	25.0	tested...
765	3.0	80.0	0.0	0.0	0.0		0.174	22.0	tested...
766	6.0	114.0	0.0	0.0	0.0		0.189	26.0	tested...
767	5.0	136.0	82.0	0.0	0.0		0.64	69.0	tested...
768	10.0	115.0	0.0	0.0	0.0		0.261	30.0	tested...

Replace Missing values with Mean Values

In preprocess tab select

Choose > filters > unsupervised > attribute > ReplaceMissingValue

Then select apply.

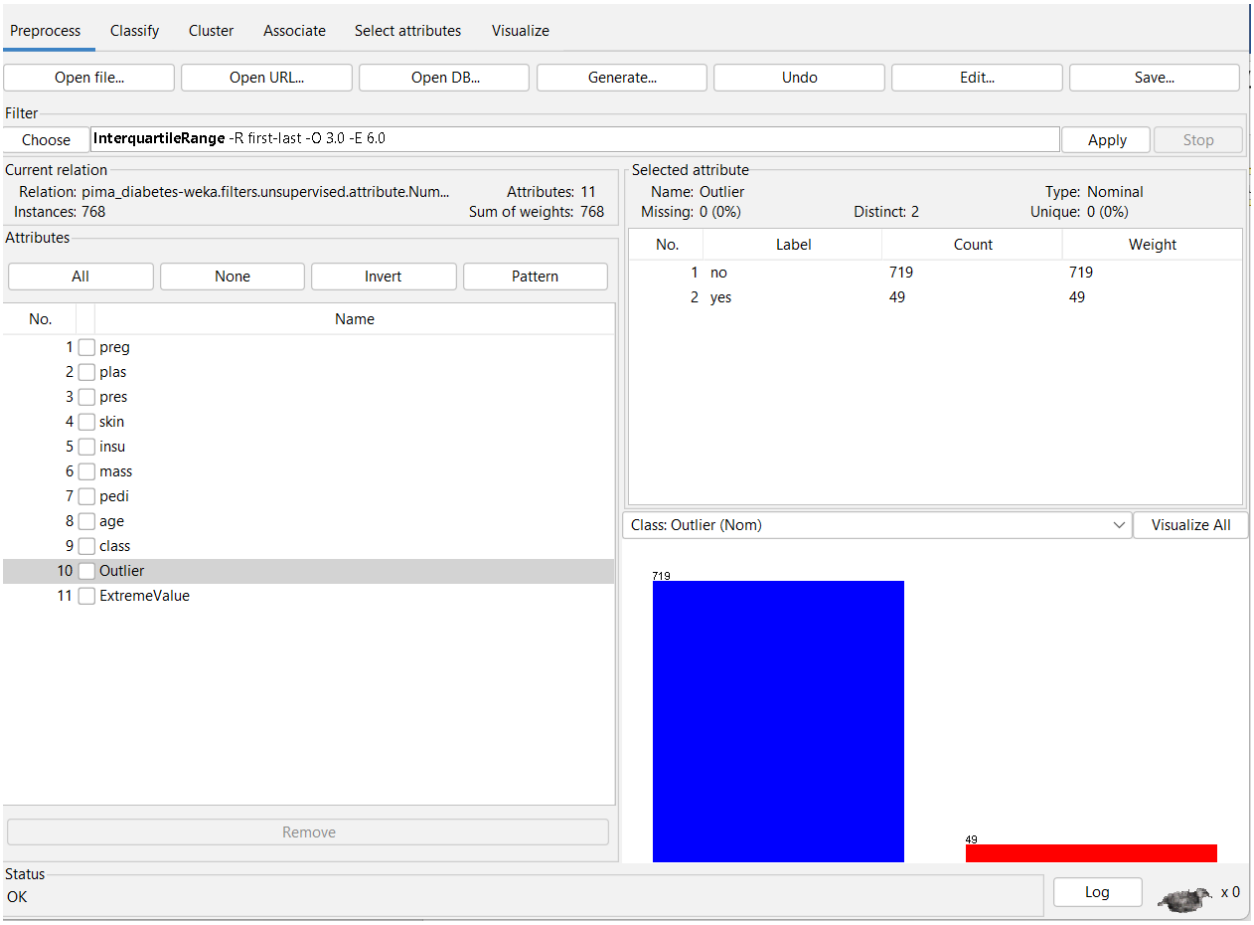
upervised.attribute.NumericCleaner-min1.0E-8-min-defaultNaN-m

ı ic	5: insu Numeric	6: mass Numeric	7: pedi Numeric	8: age Numeric	9: **class** Nominal
.0	0.0	32.4	0.2	63.0	tested...
.0	155.0	32.4	0.262	37.0	tested...
.0	190.0	32.4	0.549	27.0	tested...
.0	0.0	32.4	0.601	27.0	tested...
.0	0.0	32.4	0.141	24.0	tested...
.0	0.0	32.4	0.393	21.0	tested...
.0	0.0	32.4	0.433	22.0	tested...
.0	0.0	32.4	0.443	45.0	tested...
.0	0.0	32.45746367239099	0.232	54.0	tested...
.0	0.0	32.45746367239099	0.305	24.0	tested...
.0	0.0	32.45746367239099	0.304	21.0	tested...
.0	0.0	32.45746367239099	0.102	22.0	tested...
.0	0.0	32.45746367239099	0.572	21.0	tested...
.0	89.0	32.45746367239099	1.731	21.0	tested...
.0	0.0	32.45746367239099	0.256	25.0	tested...
.0	0.0	32.45746367239099	0.174	22.0	tested...
.0	0.0	32.45746367239099	0.189	26.0	tested...
.0	0.0	32.45746367239099	0.64	69.0	tested...
.0	0.0	32.45746367239099	0.261	30.0	tested...
.0	0.0	32.5	0.855	38.0	tested...
.0	63.0	32.5	0.318	22.0	tested...
.0	0.0	32.5	0.256	22.0	tested...
.0	0.0	32.5	0.27	39.0	tested...

Step-5: Handling Outliers and Extreme Values

Select Choose > filters > Unsupervised > attribute > InterquartileRange

After apply our dataset have 2 more attribute. Outliers and Extreme Values. Now we have to remove the outliers.

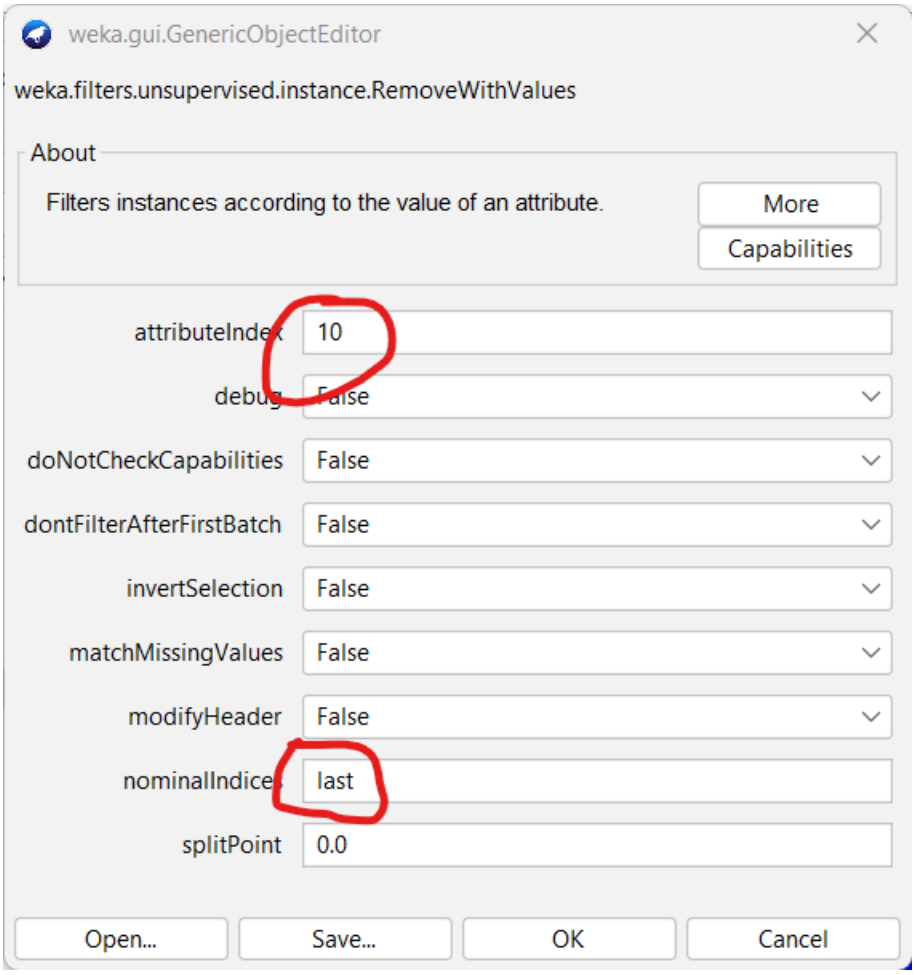


To remove the outliers

Coose > Filters > unsupervised > instance > RemoveWithValues

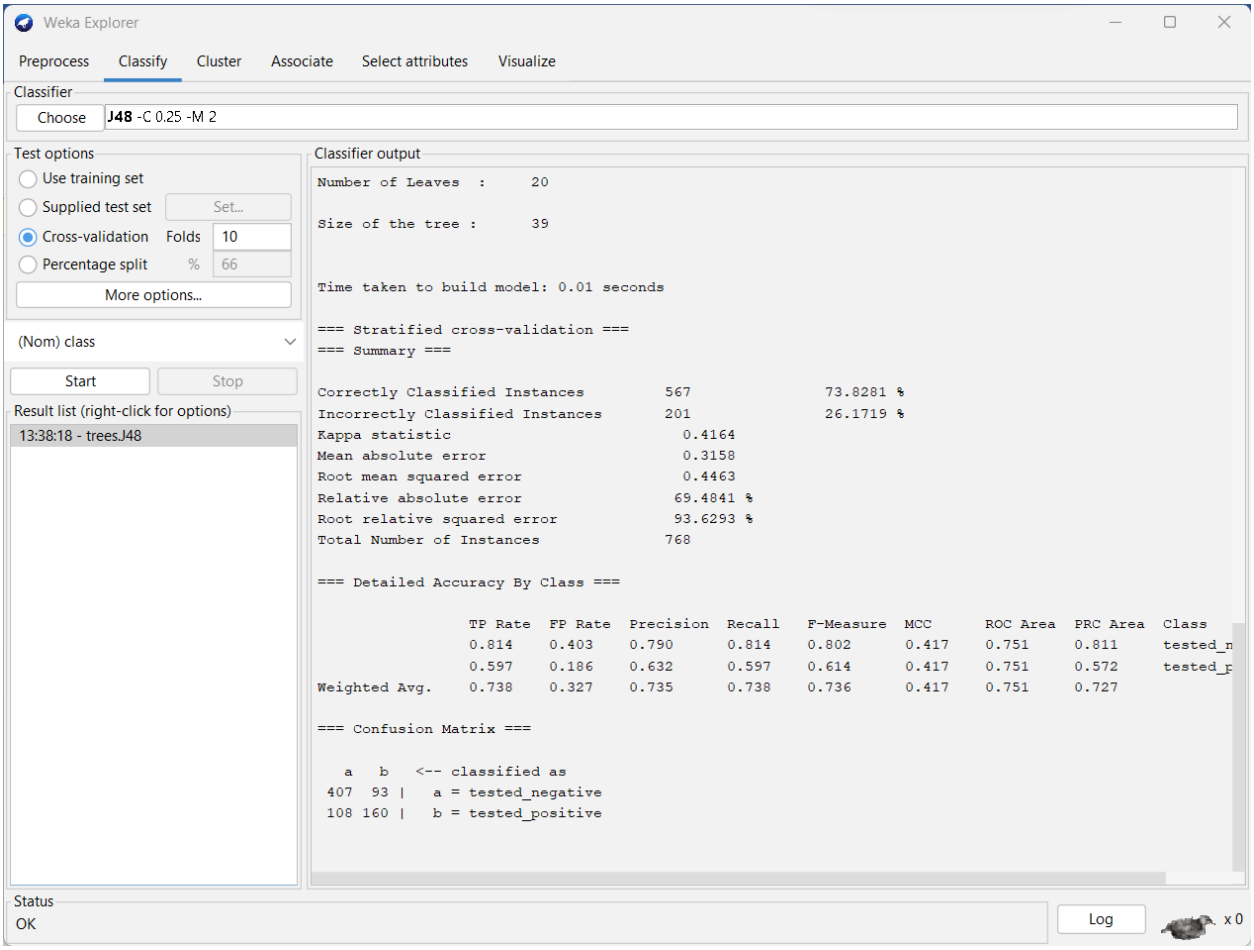
In the option of our filter e specify attribute index 10 as outliers and set nominal index = last.

After apply all outliers are removed.

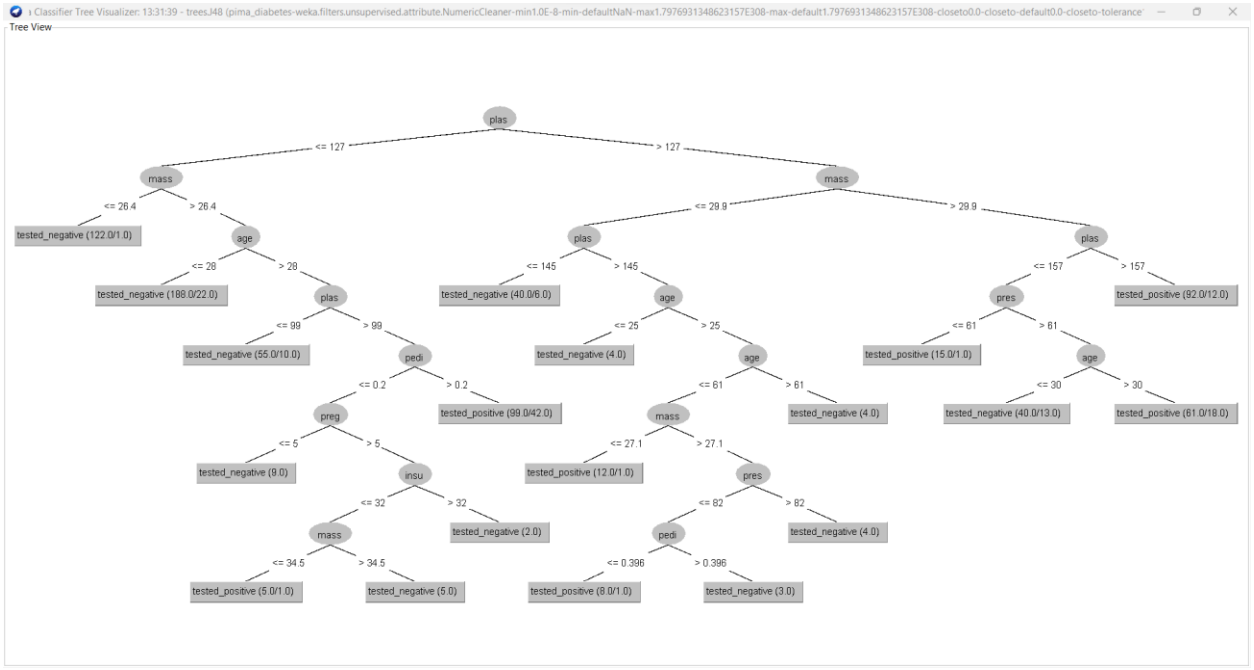


Step-6: Classification with weka

In the classification tab we we select j48 algorithm from tree category. Then start classify.



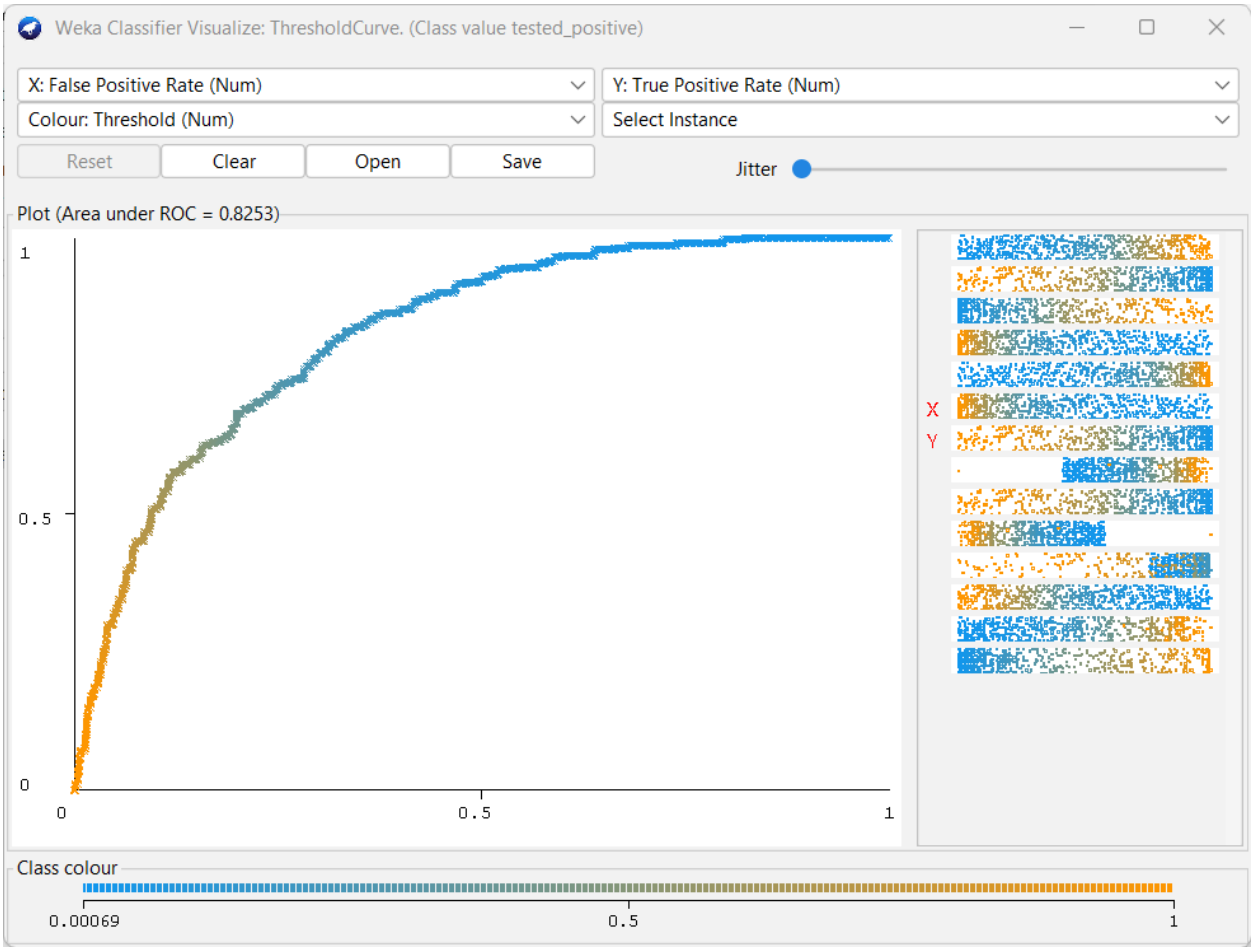
Then we visualize the tree



Step-7: Visualize ROC curve

We classify our dataset using NaiveBayes Classifier the visualize threshold curve.

Tested positive curve



Precision vs Recall curve